Solution 1: Entropy in Binary Classification

(a) By definition, the entropy of a *Bernoulli* random variable is :

$$\begin{aligned} H(\epsilon) &= -\sum_{\epsilon \in \Omega_\epsilon} \mathbb{P}(\epsilon) log(\mathbb{P}(\epsilon)) \\ &= -\theta_\epsilon log(\theta_\epsilon) - (1-\theta_\epsilon) log(1-\theta_\epsilon) \end{aligned}$$

(b) The conditional entropy quantifies the uncertainty of y if the outcome of x is given. It is defined as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable.

$$\begin{aligned} H(y|x) = \mathbb{E}_x[H(y|x=x)] &= \sum_{x \in \Omega_x} \mathbb{P}(x=x) H(y|x=x) \\ &= \theta_x H(y|x=1) + (1-\theta_x) H(y|x=0) \end{aligned}$$

Let's think about what happens with y when the values of x are given.

- If $x = 1$, the maximum between ϵ and x will always be 1 and y will always be 1. As there is no uncertainty, the conditional entropy is 0.
- If $x = 0$, the maximum between ϵ and x has uncertainty, as it can be 0 or 1. The uncertainty is given only by ϵ, because x is not random in the conditional entropy.

$$\begin{aligned} H(y|x) &= \theta_x \underbrace{H(y|x=1)}_{=0} + (1-\theta_x) \underbrace{H(y|x=0)}_{=H(\epsilon)} \\ &= (1-\theta_x) H(\epsilon) \end{aligned}$$

(c) Using the chain rule for entropy:

$$\begin{aligned} H(y,x) &= H(x) + H(y|x) \\ &= -\theta_x log(\theta_x) - (1-\theta_x)(log(1-\theta_x) - H(\epsilon)) \end{aligned}$$

(d) Now ϵ has a deterministic relation with x. Let's see how our results change with this modification.

- As ϵ and x have a deterministic relation, the uncertainty introduced by the ψ function is 0. The entropy of ϵ is given only by the entropy of x

$$H(\epsilon) = H(x)$$

- y is a function of x and ϵ, and now ϵ has a deterministic relation with x. Accordingly, we can conclude that y is now a function of x.

$$y = 2\ max\{x, \psi(x)\} - 1$$

If the value of x is given, there is no uncertainty associated.

$$H(y|x) = 0$$

- Using the chain rule for the entropy:

$$\begin{aligned} H(y, x) &= H(x) + \underbrace{H(y|x)}_{=0} \\ &= H(x) \end{aligned}$$

This result is expected. As x and y have a deterministic relation, the uncertainty is only given by x.